1. (d)
2. (a)
3. (a)
4. (c)
5. (c)
6. (a)
7. (c)
8. (b)
9. (b)
10. **BOX PLOT AND HISTOGRAM**
    A box plot is a data display that draws a box over a number line to show the interquartile range of the data. The 'whiskers' of a box plot show the least and greatest values in the data set.
    Histograms are a special kind of bar graph that shows a bar for a range of data values instead of a single value.
11. **METRICS SELECTION**
    Good metrics are important to your company growth and objectives. Your key metrics should always be closely tied to your primary objective. A good metric example might be month-on-month revenue growth or LTV: CAC ratio. 'Important' is somewhat subjective since growth for one company may be centred around revenue while another company may focus more on user growth. The key point is to choose metrics that clearly indicate where you are now in relation to your goals.
    Good metrics can be improved. Good metrics measure progress, which means there needs to be room for improvement. For example, reducing churn by 0.8% or increasing your activation rate by 3%. One exception to this might be customer satisfaction - if you're already at 100%, your team will be focused on maintaining that level instead of improving it.
    Good metrics inspire action. When your metrics are important and can be improved, you and your team will immediately know what to do or what questions to ask. For example, why has our conversion rate dropped? Did we make site changes or test a new acquisition channel? Why is churn increasing? By asking questions you can determine possible causes and work to resolve them right away.
12. **STATISTICAL SIGNIFICANCE OF INSIGHT**
    To assess statistical significance, you would use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Second, you'd calculate the p-value, which is the likelihood of getting the test's observed findings if the null hypothesis is true. Finally, you would select the threshold of significance (alpha) and reject the null hypothesis if the p-value is smaller than the alpha — in other words, the result is statistically significant.
13. **DATA WHICH DOESN'T HAVE GAUSSIAN DISTRIBUTION NOR LOG-NORMAL**
    Many random variables have distributions that are asymptotically Gaussian but may be significantly non-Gaussian for small numbers. For example, the Poisson Distribution, which describes (among other things) the number of unlikely events occurring after providing a sufficient opportunity for a few events to occur. It is pretty non-Gaussian unless the mean number of events is very large. The mathematical form of the distribution is still Poisson, but a histogram of the number of events after many trials with a large average number of events eventually looks fairly Gaussian.
    For me, the best examples come from my field of research (astrophysical data analysis). For example, something that comes up all the time is that we detect stars in astronomical images and solve for their celestial coordinates. My current project uses images about 1.5 degrees on a side and typically detects 60 to 80 thousand stars per image, with the number well modeled as a Poisson Distribution, assuming that the image is not of a star cluster surrounded by mostly empty space. That's about 8 or 9 stars per square arcminute. If we cut out "postage stamps "from the image that are half an arcminute per side, then the mean number of detected stars in them is about 2. If we do that for (say) 1000 postage stamps and make a histogram of the number

of detected stars in them, it will not look very Gaussian, but as we increase the size of the postage stamps, it becomes asymptotically Gaussian.

What generally never becomes Gaussian, however, is the Uniform Distribution. A histogram of the stars' right ascensions or declinations (the azimuthal and elevation angles used in astronomy) looks a lot like a step function, i.e., flat within the image boundaries. The positions are not uniformly spaced, but they are distributed in the same way as a uniformly distributed random variable for any size postage stamp, including the entire image.

Another example is the location of the centres of raindrop ripples on a pond; they are not uniformly spaced in (say) the east-west direction, but they are uniformly distributed.

The simplest example is the distribution of numbers that show up on the top of a fair die after a large number of throws. Each number from 1 to 6 will occur with approximately equal frequency. Increasing the number of throws will not tend to produce a bell-shaped histogram, in fact the fractional occurrence will approach a constant 1/6 over the possible numbers.

14. **MEDIAN – A BETTER MEASURE THAN MEAN**

Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

The mean overestimates where most household incomes fall.

15. **LIKELIHOOD**

The likelihood is the probability that a particular outcome is observed when the true value of the parameter is, equivalent to the probability mass on it is not a probability density over the parameter. The likelihood should not be confused with which is the posterior probability of given the data.